Ⅱ−184　零交さ波による音韻識別について

石　田　義　久
小　川　康　男

On a Speech Recognition System Using Zero-crossing Waves

Yoshihisa ISHIDA
Yasuo OGAWA

Abstract

This paper discusses about the voice recoginition system to which zero-crossing waves are applied. The device can extact both formant-like information and pitch frequency as the feature arising from zero-crossing waves, while the recognition of phoneme is obtained through a data processor. Phoneme, at the shortest from the distance between the vectors showing the normal patern, based on formant-like information, and those showing the unknown input, can be discriminated.

This data processor has a capacity of calculation as good as Digital Computer, and it can, too, memorize the program necessary to discriminate.

This study, directing towards deaf-and-dumb persons, is only a part concerning speech synthesizer and recognition system ; the problems of circuit composition and real time recognition are the chief purpose of our study.

1　まえがき

機械による音声認識の研究が，多方面で行なわれているが，その研究がいま一つ，発展性を欠く要因として，ホルマント周波数やピッチ周波数などの基本的特徴抽出さえも，なかなか容易でないという点があげられる。しかしながら，最近，これらの特徴抽出について，合成による分析，短時間スペクトル分析などの優れた手法を用いた研究が発表され，この結果，90%を越える高い識別率をもつ認識装置が実現されるようになった[1]。しかし，これらの装置はいずれも，複雑な計算処理を必要とし，このため，実時間の認識を行なうには，処理速度の速い計算機が必要となる。

さて，本研究は，ろう唖者を対象とした音声合成．認識系の研究のうち，とくに認識系に関する研究の一部をなすものである。そこで，本文で述べる装置は，連続音声の識別を最終目的としているが，ろう唖者を対象とすることから，まず研究の初段階として，文章を話者に一語一語明瞭に発声させ，これによって認識装置の簡単化を図ることにした。

一方，ろう教育の現状からみると，ろう者の発声訓練においては，このような識別装置より，むしろ音声の特徴抽出がより重要であると考えられている。すなわち，ろう者の発声訓練において，正常者との発声の比較はなかなか困難であり，このため正常者の発声音を適当な装置によってパターン化し，これとろう者の発声を比較せしめて，発声者の矯正を計ろうとするものである。

筆者らは，これまで音声の特徴抽出に始まり，閾値論理回路による音声識別装置と零交さ波分析装置とについて研究を進めてきた[2][3]。そこで，現在は，過去数年にわたる基礎実験をもとに，ろう者の発声訓練が可能な音声認識装置の試作を進めている。本研究は，このうち零交さ波を用いた特徴抽出回路の試作がほぼ完成したので，その詳細を述べるものである。

零交さ波を利用した分析装置は，前号に述べたように[4]，広帯域のフィルタを使用できるため。過渡音やわたりなどに対して，その時間的変化を忠実に再現できるという特徴をもっている。一方，定常母音の識別に必要な情報として，第1ホルマント周波数と第2ホ

ルマント周波数に，それぞれに対応する２つのローカルピーク情報をとり上げ，以後の処理の簡単化を図っている。また，本装置では，ローカルピーク以外の情報として，話者のピッチ周波数を導入し，個人性による影響を考慮している。

2　音韻識別に必要な情報量

定常母音の識別に必要な情報量として，第１，第２ホルマント周波数があげられる。しかし，第１，第２ホルマントを直交軸とする従来のパターン識別では，認識対象となる話者の数が増えると，その個人性に起因して，母音相互間に重なり合いを生じ，判別できなくなる場合がある。そこで，本装置では，第１，第２ホルマント以外の情報量として，ピッチ周波数の抽出を行ない，個人性による影響について考慮した。

一方，破裂子音（b，d，g，p，t，k）と鼻音（m，n）は，ホルマント周波数の時間的変化によって，特徴づけられることが，ハスキンス研究所における音声合成実験により指摘されている[5]。

声道内のせばめ（調音点）の付近に発生する乱流を音源とする無声摩擦音（s）は，子音部における周波数成分によって特徴づけられる。すなわち，摩擦音は一種の白色雑音を音源としているため，他の音韻に比べると，子音部における高周波成分の強さが顕著となる。

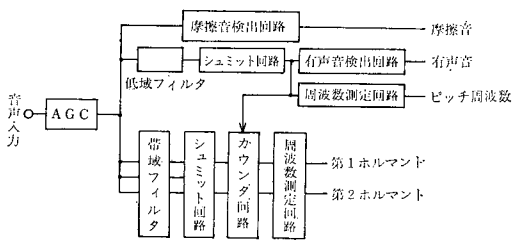
有声摩擦音（z）は，子音部においてピッチ波形が観測されるほか，無声摩擦音と同様な特徴をもっている。

声門子音（h，f）は，無声摩擦音同様，子音部における周波数成分によって特徴づけられる。ただし，声門子音の場合は，低い周波数領域にエネルギーが集中している。

このように，各音韻は，第１，第２ホルマント周波数の時間的変化，ピッチ波形の有無，子音部における高周波成分等により特徴づけられる。

3　特徴抽出部の回路構成

第１図に，入力音声からローカルピーク（認識系からみたホルマント周波数は，ローカルピークと呼ばれる[6]）やピッチ周波数等の音韻識別に必要とする情報をとり出す，いわゆる特徴抽出部の回路構成を示す。各部について動作原理を述べれば，およそつぎのようである。



第１図　特徴抽出部の構成

3—1　自動利得調整回路（ＡＧＣ回路）

音韻識別に対し，二次的な情報とされている音圧レベルの正規化を行なうため，アナログ割算器を用いたＡＧＣ回路が用いられている。ＡＧＣ回路の制御電圧は，全波整流回路と低域フィルタから成る音声の包絡線検出回路の出力によって与えられる。

3—2　低域フィルタ

入力音声からピッチ成分をとり出すためのフィルタで，その出力はシュミット回路によって波形整形される。パルスに変換されたピッチ波形は，特徴抽出部を同期制御するもので，有声部の検出にも使用される。

3—3　摩擦音検出回路

サ行やザ行のような摩擦音は，5KHz 付近のエネルギーが，ある一定値より大きくなることにより検出される。摩擦音検出回路の出力は，発声される音韻が摩擦子音であるときに“１”，それ以外は“０”となる。

3—4　帯域フィルタ群

音声を零交さ波に変換する場合，同一帯域内に，ホルマントのようなエネルギーの大きな成分が２つ以上あると，混変調によって，零交さ間隔が，必ずしもホルマント周波数に対応しない，という現象が生ずる[4]。そのため，本装置では，音声周波数帯域を３つに分け，混変調による影響を軽減している。

帯域フィルタの出力は，シュミット回路によって零交さ波に変換され，ローカルピークを抽出するためのカウンタ回路に入力される。

3—5　カウンタ回路

カウンタ回路は，零交さ波からローカルピークを抽出するために使用されるもので，第１ホルマント領域において，３回の零交さに要する時間が，さらに第２ホルマント領域において，８回の零交さに要する時間が，それぞれ測定される。

3—6　周波数測定回路

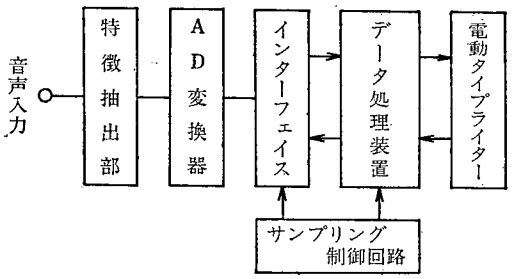
カウンタ回路の出力は，アナログ掛算器と積分器とから成る周波数測定回路[7]によって，アナログ電圧に変換される。一方，ピッチ波形も周波数測定回路により，周波数に比例したアナログ量に変換され，データ処理装置へ送られる。

4　音韻認識部

特徴抽出部によって得られたローカルピーク，ピッ

チ周波数等の情報は，10msec 毎にサンプリングされ，音韻決定回路に加わる。音韻決定回路は，ディジタルシステムから成る一種のデータ処理装置で，あらかじめプログラムしたアルゴリズムにしたがって，音韻の識別を行なう。

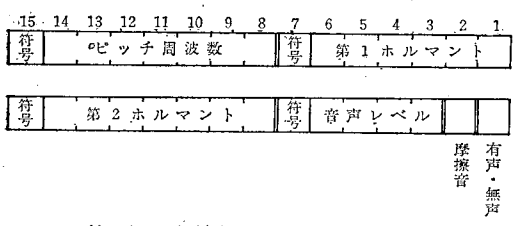
データ処理装置への入力は，すべて“0”，“1”の論理レベルによって行なわれ，認識結果は電動タイプライタからタイプアウトされる。このため，現在は，認識結果をタイプするまで比較的時間を要するが，ランプ表示，ブラウン管表示等を用いれば，容易に実時間処理が可能となる。データ処理装置および周辺回路の構成図を第2図に示す。ここで，A/D 変換器は，特徴抽出部によって得られたローカルピークとピッチ周波数の，それぞれに比例したアナログ量を2進符号に変換する。これに加えて，音圧レベルをも2進符号に変換する働らきをもっている。



第2図　データ処理装置と周辺装置の構成

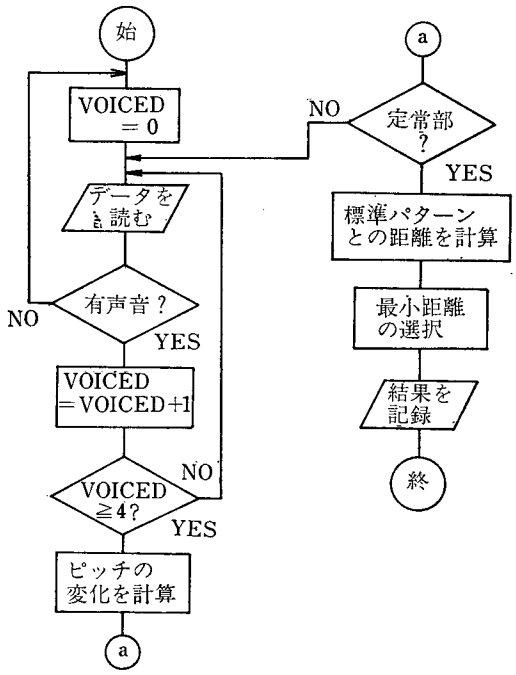
A/D 変換器によってディジタル信号に変換された特徴量は，10msec 毎にデータ処理装置に入力され，記憶部に順次蓄積される。記憶部は，16ビット1ワードで8Kワードの容量をもち，特徴抽出部によって得られたデータを記録するとともに，認識に必要なプログラムが内蔵されている。さらに，記憶部には，150 Kワードの記憶容量をもつカセット式メモリーが外装されており，連続音声認識装置にも対処できるようになっている。

記憶部において，ローカルピーク，ピッチ周波数等のデータは，第3図に示すようにビット配分され，各サンプリングタイム毎に，順次記憶される。



第3図　記憶部におけるビット配分

第4図は，単母音を識別するためのプログラム流れ



第4図　母音識別の流れ図

図である。同図において，ピッチ周波数の変化の計算は，ピッチの動きが小さく，母音がほぼ定常状態にあると見なせるか，否かの判定に使用される。そこで，母音の決定は，音声がほぼ定常部にあると見なせる時点においてのみ行なわれる。

ピッチ周波数の変化 Δf_0 は，次式に示すように最小自乗法による近似的傾きによって計算される。

$$\Delta f_0=\frac{1}{2}[f_0(t)-f_0(t-3)]+\frac{1}{4}[f_0(t-1)-f_0(t-2)]$$

上式において，$f_0(t-n)$，$n=0\sim3$ は，ある時刻 t におけるピッチ周波数 $f(t)$ と，それより n サンプル前のピッチ周波数を表わす。

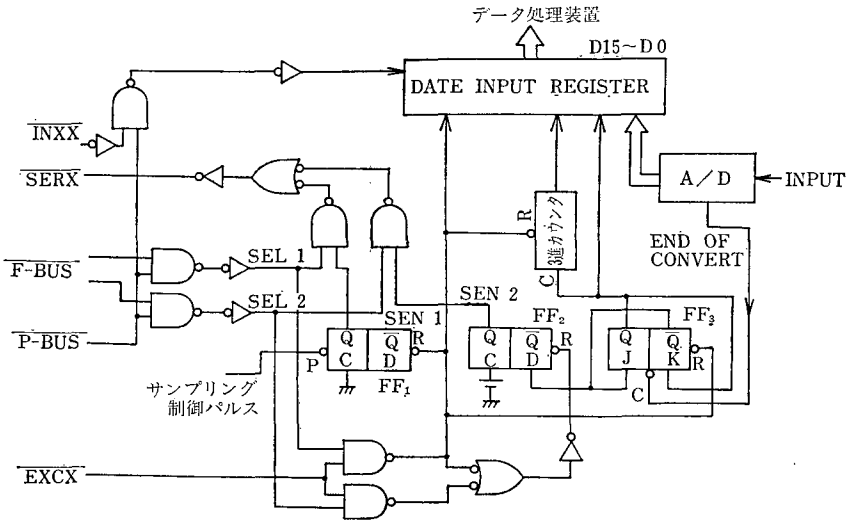
母音の識別は，未知パターン $x=(x_1, x_2)$ と標準パターン $y_i=(y_{i1}, y_{i2})$ との間の次式で定義される距離 $d(x, y_i)$ を求めることによって行なわれ，最小距離にある音韻が認識される。

$$d(x, y_i)=\sqrt{\sum_{n=1}^{2}(x_n-y_{in})^2},\quad i=1, 2, \cdots, 5$$

ここで，(x_1, x_2) は未知音声の2つのホルマント情報を，(y_{i1}, y_{i2}) は5母音のうち，i 番目の母音の標準的なホルマント情報を表わす。

5　データ処理装置の周辺装置

A/D 変換器とデータ処理装置との接続回路につい



第5図　インターフェイス

て述べると，およそつぎのようである。第5図は，A/D 変換器に対するインターフェイスで，まず，SELECT 命令1により，入出力装置の選択信号 P-BUS (Peripheral Address Bus) と F-BUS (Function Bus) とが指定される。このとき，リセット信号 $\overline{\text{EXCX}}$（ここで——，は負論理を意味する）によって DATA INPUT REGISTER（以下略して，D. I. R. とする）および他のすべてのフリップフロップ（以下略して，FF とする）がクリアされ，データ待ちの状態となる。

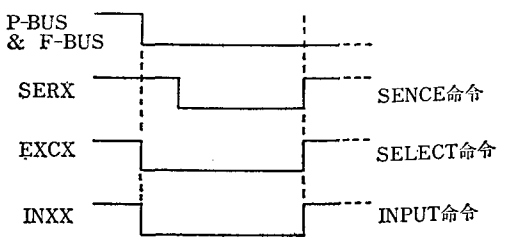
つぎに，SENCE 命令1により，FF_1 の状態が確認され，もし High level にあれば（サンプリング制御パルスが入力されると，High level になる），センス信号 $\overline{\text{SERX}}$ が出力され，DATA READY の状態となる。この状態から，AD 変換が開始されるわけであるが，変換が終了すると，END OF CONVERT 信号により，FF_3 の状態が反転し，D. I. R. の左半分 (D15~D8) にデータが入力される。引続き，AD 変換されたデータが D7~D0 に入力されると FF_2 の出力状態が反転し，SENCE 命令2によって，処理装置本体へセンス応答が送り込まれる。このセンス応答は，データが D. I. R. に完全にセットされているか否かを確認する信号である。

データが D. I. R. にセットされると，INPUT 命令（この命令が実行されると，P-BUS と F-BUS により入力装置が選択され，同時に入力信号 INXX が Low level となる）によって，D. I. R. の内容が，16ビット並列に処理装置内のレジスタに転送され，記憶部にストアされる。

データが処理装置に転送されると，SELECT 命令2により，FF_2 がクリアされ，次のデータ待ちとなる。同様な順序で D15~D8 がセットされると，カウンタ回路の出力は High level となり D1, D0 の出力は，処理装置と切り離される。これは，有声／無声の区別および摩擦部の有無を表わす2ビットのディジタル信号を記録するためのものである。

以上の操作が順次繰返され，音韻の識別に必要なデータが記憶部にストアされると，SELECT 命令1により，FF_1 の出力が Low level となり，処理装置への入力が停止する。

さて，各命令に対する時間関係を示せば第6図のようである。SENCE 命令，SELECT 命令，および INPUT 命令のいずれかが実行されると，P-BUS と F-BUS はそれぞれ Low level となり，入出力装置が選択される。



第6図　タイミングチャート

SENCE 命令は，処理装置からセンス信号を出して，その応答があったとき（SERX が Low level になる）に，次の命令を実行するものである。

SELECT 命令が実行されると，EXCX 信号によっ

て，バッファレジスタがクリアされ，入出力装置が指定される。

INPUT 命令は，P-BUS と F-BUS によって指定された入出力装置から，データ処理装置内のレジスタにデータを転送するもので，INXX 信号によって行なわれる。

6 ろう教育への応用

ろう者の発声訓練において，正常者の音声を何らかの装置によって視覚化することは，音韻の識別よりも，むしろ重要であると考えられている。すなわち，正常者の音声を視覚表示して，これと，ろう者の発声する音声を比較せしめ，発声音の矯正を行なおうとするものである。

本研究は，このようなろう者の発声指導が一つの目標であり，つぎに示す特徴をもっている。従来の研究と併せて，ピッチ周波数の変化を実時間でブラウン管にプロットできるため，ろう者にアクセント指導を行なうことができる。

音韻識別に重要なホルマント周波数を，比較的簡単な回路構成によって視覚表示できるため，ろう者の発音訓練を効果的に進めることができる。

一方，別稿に述べる閾値論理回路による音声識別装置は，高い識別率（5母音に対し，ほぼ100％）をもっているので，ろう者の発声訓練が適切に行なわれているかどうかの検討に有効である。

7 むすび

本研究は，零交さ波を用いた分析装置において，その主要部である特徴抽出部とデータ処理装置とについて述べたものである。

ろう者の発声訓練において，音声の特徴抽出は重要な問題となっている。筆者らは，ろう教育に対し，閾値論理回路による音声識別装置と併せて，本装置の活用を進める予定である。

終りに，直接御指導を賜っている本学助教授本多高先生に感謝の意を表する。また，日頃御指導御鞭撻を賜っている本学教授後藤以紀先生，同西山栄技先生，同助教授天野正章先生に感謝の意を表する。さらに，データ処理装置の開発，製作に御協力下さった亜細亜製作所（株）電子計算機課海老沢繁一氏をはじめ，同課の方々に深甚なる謝意を表する。また，本研究に御協力下さったゼミナール学生諸氏に謝意を表する。なお，本研究は，文部省科学研究費によることを付記する。

参考文献

(1) 角川，中田："「合成による分析法」によるフォルマント周波数の抽出" 音響学会誌，20，1-13(1964)

(2) 小川，森野他："閾値論理回路による音声認識について" 音響学会講演論文集，2-2-13 (1972-10)

(3) 石田，小川他："ホルマント周波数検出装置と閾値論理回路による音声認識" アナログ研究会資料，VOL. 12，No. 7 (1972-9)

(4) 石田，小川："零交さ間隔によるホルマント周波数の検出法について" 明大工学部研究報告，No.26-27，II-172 (1973)

(5) 越川，中田他："聴覚と音声" 電子通信学会

(6) 松岡，城戸："連続音声のスペクトルのローカルピークの軌跡について" 音響学会講演論文集，2-2-6 (1972-10)

(7) 石田，小川："ピッチ抽出装置について" 明大工学部研究報告，No.26-27，II-163 (1973)